

学力テストにおける妥当性の吟味の必要性

正 田 良

1. はじめに

東京都が公立小学校の5年生に行った学力調査のうち2007年度のものとは2008年度のものについて、その妥当性の検討が本研究の内容である。10年前のテストについて、あまり建設的な提言のないものを書くのは、さもしいと、躊躇しないこともない。しかし、この10年の間にその問題点が改善されるかと言えば、調査問題の形式が変わっても、学力調査の結果をどう扱うかという根底的な問題に関しては、改善されるどころか、こうした政治的な構造はむしろ定着した感もあるほどである。そこで、この10年間の傾向を振り返るとともに、具体的なデータを持っているほぼ10年前のものの検討を行うものである。

全国学力・学習状況調査は、21世紀に入ってからでは、2007年度から小学校6年生、中学校3年生を対象に4月に行われている。これの前史として「全国中学校一斉学力調査」が、1956年から1966年にかけて行われている。この当時は、地域間の競争激化が問題となり、教職員組合でも「学テ闘争」として大きな問題として取り上げられた。しかし、2007年度からのものは、その間に自民党と民主党との間の政権のやりとりがあり、悉皆調査か標本調査かの若干の変動はあったものの、異議申し立ては目立ってはいない。むしろ、各都道府県では、小学校5年生や中学校2年生を対象とする独自の学力調査を行って、既定の事実としてそれへの対応を行っている。上述の東京都の調査も、この独自の学力調査の一例と位置付けられるものである。

東京都教育委員会（2016）の「はじめに」に、そのような経緯が説明されている。

東京都教育委員会は、国に先駆け、中学校では平成15年度、小学校では平成16年度から「児童・生徒の学力向上を図るための調査」を開始しました。この間、「教科別・観点別の学力定着状況調査」、「基礎的・基本的な事項に関する調査」、「問題解決能力等に関する調査」を実施し、…（後略）

つまり東京都は小学校5年生と中学校2年生に、平成16年（2004）から、調査を行っており、そのテーマはその時点で問題となっている表1の課題が選んでいる。ここで検討する、「問題解決能力等に関する調査」は、現在では行われてはいない。「読み解く力」に関する問題を加味した国算社理の4教科のテストへ、発展的解消をしている。

表 1 東京都の「学力向上を図るための調査」での課題

期間	課題
平成 18 年（2006 年）度	「問題解決能力等に関する調査」 「教科別・観点別の学力定着状況」
平成 19 年（2007 年）度	「問題解決能力等に関する調査」 「基礎的・基本的な事項に関する調査」
平成 20 年（2008 年）度	「問題解決能力等に関する調査」 「基礎的・基本的な事項に関する調査」
平成 21 年（2009 年）度	「基礎的・基本的な事項に関する調査」
平成 22 年（2010 年）度	「読み解く力に関する調査」 「基礎的・基本的な事項に関する調査」
平成 23 年（2011 年）度	「読み解く力に関する調査」
平成 24 年（2012 年）度	「読み解く力」の定着状況と、国算社理に関する目標や内容の実現状況
平成 25 年（2013 年）度	
平成 26 年（2014 年）度	
平成 27 年（2015 年）度	
平成 28 年（2016 年）度	

※ 『平成 28 年度 児童・生徒の学力向上を図るための調査 報告書』などは、それぞれ下記サイトからリンクがある。

http://www.kyoiku.metro.tokyo.jp/buka/shidou/chosa_top.htm

2. 問題解決能力等に関する調査の概要と合成質問紙の作成

「問題解決能力等」に関する質問紙調査（平成 18 年度～ 21 年度）の目的は、

(1) 児童・生徒一人一人に確かな学力の定着と伸長を図る。

(2) 各教科の目標や内容の実現状況を把握し、指導方法の改善・充実に生かす。

となっている。そのコストなどの評価を度外視して考えるとして、この趣旨自体は結構なことのように思える。

この対象学年は、全国学力テストの対象となっている小 6 と中 3 のちょうど前の学年。B 問題として「活用」に関する学力が含まれるこの調査結果と、「問題解決等」に関する学力との関連を調べようとしていることは容易に想像できる。小 5 を対象とする質問紙に注目する。「問題解決能力」に関して 5 つの観点

ア：問題を発見する力

イ：見通す力

ウ：適応・応用する力

エ：意思決定する力

オ：表現する力

を設け、それぞれの観点に対応する問題を含む 8 題で質問紙が構成されている。

表 2：2つの年度の小 5 質問紙での 5 つの観点

年度\問	1	2	3	4	5	6	7	8	9
2007 年度	ア	ウ	ウ	ウ	エ	ウ	イ	オ	－
2008 年度	ア	イ	ウ	ウ	ウ	エ	オ	ウ	－
今回	08-1	08-2	08-3	07-4	08-5	08-6	07-7	07-5	08-8
	ア	イ	ウ	ウ	ウ	エ	イ	エ	ウ

つまり、いわゆる「問題解決型授業」の様式と、それを現実世界に「活用」する過程を意識した順番とみることができる。表 2 に、2007 年度並びに 2008 年度に実施された調査について、東京都の調査報告書によって、それぞれの問題がア～オの 5 つの観点のどれを測ることを意図したかを記した。

これらの合計 16 問のうちから、9 問を選んで、大学生を被験者とする質問紙を作った。表 2 の「今回」がそれで、上の行には、「08-1」（2008 年度の第 1 問の意味）などとして元の東京都の問題のどれにあたるか、下の行には、東京都がどの観点について測るための問いとしているのかを記した。

東京都の問題は、ウが 4 題で、他の観点は 1 題ずつという構成であったが、合成して作った問題では、ア・オ以外の 3 つの観点については複数の問題が含まれるようにした。オが 0 題、アが 1 題とこの 2 つの観点については犠牲にせざるを得なくなった。必ずしも理想的なものとは言えないが、後に述べるように被験者として大学生に協力を依頼したので、彼らの我慢の程度を考えて問題数を限らざるを得なかった。ウがほぼ半数であるという特徴を継承し、また、解くにあたっての面白さも若干考慮した。

3. 調査の結果

2009 年 10 月並びに 2010 年 4 月に、複数の大学の初等教員養成課程の学生(計 234 名)を被験者に、前節の質問紙調査を実施した。小学生へのテスト時間は 8 題を 40 分で回答するものであったが、9 題を 20 分で回答させた。被験者は時間に不足を感じては居なかった。

小学生のための問題を大学生が答えるのだから全部できてしまい関連に関する情報が得られないのではないかと心配をしたが、意外と勘違いをする人が多かった。表 3 に今回の正答率と、元になった東京都の問題が実施された当時の小学生の正答率（東京都の両年度の調査報告書による）を記す。また、表 4 に各問について、正答を 1、誤答を 0 とするダミー変数とした場合の、各問間の相関係数、並びに、「予想者数」を記す。予

想者数とは、一番相関が高いのはどれとどれかを被験者がどう予想したかの人数である。必ずしも全員がこの予想をしているわけではない。

残念ながら本物の調査に比べると被験者の人数は多くはない。そこで十分な議論はできないが、この結果から内的整合性についての議論を試みよう。

表 3：小学生と大学生の正答率の比較

被験者\問	1	2	3	4	5	6	7	8	9
小学生	08-1	08-2	08-3	07-4	08-5	08-6	07-7	07-5	08-8
	0.82	0.60	0.58	0.31	0.48	0.51	0.55	0.58	0.65
大学生	0.88	0.90	0.78	0.68	0.88	0.91	0.87	0.82	0.94

表 4：大学生を被験者とした今回の調査の基本統計量

問題	1	2	3	4	5	6	7	8	9	
正答	3	2	4	3	2	3	1	4	4	
正答者数	205	210	183	160	207	214	203	193	219	
全回答者	234	234	234	234	234	234	234	234	234	
正答率	88	90	78	68	88	91	87	82	94	
相関	1	2	3	4	5	6	7	8	9	予想者数
1		0	1	1	1	4	2	0	3	1
2	0.09		1	0	3	28	2	4	7	2
3	0.02	0.13		2	1	5	10	9	1	3
4	0.00	0.01	-0.09		11	2	0	6	1	4
5	0.27	0.10	0.00	-0.02		0	0	5	1	5
6	0.07	0.15	-0.01	0.02	0.08		6	75	2	6
7	-0.11	0.08	0.01	0.14	0.25	0.02		0	0	7
8	0.00	0.14	-0.11	0.07	0.22	0.14	0.15		0	8
9	0.01	0.20	0.07	-0.03	0.23	0.11	0.21	0.20		-

相関係数の最大は問 1 と問 5 の間の 0.27 であった。相関が全般的に大きいとは言えない。そこで敢えて続きを行うのもためられるが、念のため、各問の構造が質問紙作成者の意図と一致しているかを調べる方法のひとつである因子分析を行ってみよう。正田 (2008b) に概説したように、因子分析とは、相関が高いものはある場合に、そのどちらにも共通する原因（因子）があると考えることで、多次元量をより少ない次元で説明しようとする手法である。相関行列を対角化し、その固有値が小さなものを誤差として無視する。その小ささの判断には、いろいろな方法があるが、多くの事例で採用されるのは、「1 未満は小さい」とする基準である。

表 4 の 234 人分のデータを、主因子法バリマックス回転による因子分析を行った結果、表 5 のようになったので、因子の数は 1 とするべきと判断する。

表 5：因子分析の結果

	固有値	寄与率	累積寄与率
因子 1	1.853	0.603	0.603
因子 2	0.678	0.221	0.824

これによって、因子負荷量を計算した結果が表 6 である。因子負荷量が 0.45 を超えるものを太字にした。ここでは、問 5・7・8・9 をひとつのグループとなった。作成者の意図で「ウ」とされた 4 問のうち 2 問はこのグループに入らなかった。また、このグループの 4 問のうちイ・ウ・エの 3 種類の作成意図が見られる結果となった。

表 6：因子分析の結果（統計ソフト Stat Partner による）

	共通性	独自因子	因子 1	因子 2	
問 1	0.160	0.840	- 0.147	0.372	ア
問 2	0.160	0.840	0.192	0.352	イ
問 3	0.073	0.927	0.038	0.268	ウ
問 4	0.255	0.745	0.251	0.439	ウ
問 5	0.487	0.513	0.673	0.184	ウ
問 6	0.142	0.858	0.051	0.373	エ
問 7	0.493	0.507	0.692	0.116	イ
問 8	0.314	0.686	0.492	0.267	エ
問 9	0.588	0.412	0.741	- 0.196	ウ

一番似ているのは何と何かという質問に対して、一番多く回答されたのが、「問 6 と問 8」である。問題を見れば、両問共に複数のイベントを時系列に位置付けるということを狙った問という共通点がある。作問意図「意思決定する力」というためには、時系列に関わらない意思決定問題についても調べる必要がある。また、正答・誤答の相関は低くなっている。

先に述べたように問 5・9 は同じ「適用・応用」の観点に関するものだが、問 5 は地理的な知識技能で、問 9 は計算の技能、あるいは計算しようとする意欲があるかどうかを聞いている。それに東京都が「見通す力」として意図した問 7 が絡む。こうなってくると、他の解釈に立たないとこの因子負荷量が示す構造を説明できない。この因子は、質問紙の後半に当たる問題に効いている。今回の問題は、9 問ともに 4 択問題だった。これは、不適当と思われる選択肢を消して行って、しかもあまり余計なことを考えないで残りのものを選ぶという一種の受験テクニックの有無が大きく効く問題形式である。9 問全部にこの要素があるが、回答の後半で時間と意欲とが少なくなったときにこうしたテクニックが効いた。つまり問題間で共通する一番大きな要素は「4 択問題であること」かもしれない。

4. 教育および教育評価に対する政治的な側面

やや余談に属する話題であるが、小泉純一郎（2001-2006 年首相在職）は、短く耳触りのよい文語的表現で、広く大衆の賛同を集めた。下記の 2 つは、そのうちの典型的なものである。

- ・聖域なき改革（数値目標をたて、その達成によって税金が有効に使われたかの説明が可能であるかどうか（accountability）を重視する一方、社会保障や福祉を削減する）
- ・官から民へ（小さな政府を標榜する。この点は公共事業による景気高揚などの古典的なケインズ理論に拠った田中角栄などの旧来の自民党の政治手法と対照的な手法であって、小泉自身「自民党をぶつつぶす」とさえ口にした）

これらに「新自由主義」の特徴を見ることができる。また、安倍晋三（2006-2007、2012- 年首相在職）の

- ・美しい国日本（伝統や道徳を重視する）

も合わせれば、「新保守主義」（渡辺治, 2013:p.109）の政治手法を表していると言える。

表 7： 新自由主義を政策の軸とする政治家

	大統領もしくは首相	任期
アメリカ	ロナルド・レーガン	1981 年 - 1989 年
イギリス	マーガレット・サッチャー	1979 年 - 1990 年
日本	中曽根康弘	1982 年 - 1987 年
	小泉純一郎	2001 年 - 2006 年

表 7 に概観するように、新自由主義の潮流は、英米に於いて、1980 年代から 1990 年代にかけて顕著になる。サッチャー政権は、ナショナル・カリキュラムとそれの各キー・ステージの最後に課せられるナショナル・アセスメントとを制度化した。アメリカでは、次に記すようなやや息の長い経緯がある。1983 年に答申された「危機に立つ国家」（A Nation At Risk）は、ジョージ・ブッシュ大統領政権（共和党）下での各州の合意を経て、さらにクリントン大統領（民主党）への政権交代の後に、「どの子も置き去りにしない法」（No Child Left Behind Act）が、2003 年度から実施される。吉良直（2009）が、「1990 年代からのアカウントビリティ重視の改革が実施される中で、学力向上が両党にとっての至上命題とな」ったと指摘するように、こうした政党を超えて、アカウントビリティ（accountability）が重視された。このように新自由主義の潮流は根強く、イギリスの政権のキャッチボールで保守党から政権を受け取った労働党のブレア首相（任期 1997 年 - 2007 年）でさえ、「ネオ・レイバー」と、保守党の「ネオ・コンサバティブ」を踏襲

する政策をとらざるを得なかったのである。

日本では、中曽根康弘の政権がこの時期に対応する。確かに、国有鉄道の分割（JR）などの政策を実行しているが、教育の分野での典型的な制度改革である全国学力・学習状況調査の再開は、小泉首相が2004年11月の経済財政諮問会議へ提案したことに端を発している。

1980年代の日本の教育政策は、むしろ「ゆとり」をスローガンとしている。当時の日本の学校教育の状況は、1970年代の高等学校での「落ちこぼし」問題、1980年代の中学校の荒れなどの、子どもからの「異議申し立て」的行動が顕著であったこと。また、産業での状況は、「80年代、アメリカやイギリスの製造業は、日本、(旧)西ドイツなどとの競争に敗退していた。…(中略)…(アメリカ政府は)日本の自動車会社に対して対米輸出自主規制と求めるという保護貿易主義をとっていた」(服部茂幸, 2013: p.39)と指摘されているように、日本は新自由主義に付随するところのグローバリゼーションの立場をアメリカに対してとれなかった。

5. 教育評価に求められるもの

他方、教育評価について、東洋(2001)は、「清く、正しく、愛深く」の3者が必要であると指摘している。前2者は信頼性と妥当性のことであるが、後者は、評価する者が、評価される者をどのアングルから撮れば善く見ることができるかを意識することである。また、評価が何の目的のために誰によって行われるべきかに関して「社会」から「個人」への系列を指摘している。正田(2008a)に述べたように、子どもに対する評価には、「それぞれの学年に応じた成長を支援するため」に、「作品創りの能力を、それぞれの学年に応じて具体的な作品の相互評価などを通じて体系的に示す」必要があるが、「社会」即ち、行政のための資料に関しても、多次元的な評価を標榜しながら、単なる各観点の合計と言う1次元的な順位による判断をするのは、教育評価としては、「愛深く」とは言えない。「愛深い」教育評価のためには、テストの結果が多次元的な情報が取り出しやすい必要がある。そしてその大前提として、各次元の内的妥当性が保障されていることが大切である。

いま日本の地方自治体では、市議会などでの審議で、税金使用が有効であったかという説明責任(accountability)が取り沙汰されている。その答弁が市長であれ教育長であれ、その資料を作ること。またしっかりとした結果を答弁できるような実績作りが、教育委員会に、学校に求められている。それは各学級のスコアが上がるような指導が熱烈

に求められる結果となる。しかし、その熱烈的な努力にふさわしい内的妥当性があるのか。大学生の正答率が小学生よりも高い分、いわば「天井効果」（数値のデータを扱う場合、平均点がそのスコアの上限に近い場合、上限を超えるスコアに対応する力を持っている被験者相互の差が、スコアとして現われなくなってしまう現象）があって、相関が低くなったのかもしれない。その譲歩をしたとしても、前節の結果は、この点に関して大いに疑問を呈するものである。

悉皆調査であるので、事前の予備調査は行いにくいだろう。しかし少なくとも、事後に同じ側面を測っている問題のスコア間の相関が高いかどうかという、「内的妥当性」が調べられ、学校現場へ報告されるべきだ。

この語句「内的妥当性」について補足的な説明を以下に記しておこう。熱烈に努力した結果は、当然乱数による数値として生ずるものであってはならない。また、その数値を分析して、その後の政策決定や行政での対応が為されるなら、数値に科学的根拠がある必要がある。しかし、同じものを測っているとされる昨年の問題のスコアと今年のそれとが、相関が無いなら、結局そのスコアは測っている概念を示すものではなく、単なる当たり外れ、つまり乱数のようなものになってしまう。これへこれまでの教育実践との関連を考えると、今後の教育の施策を考えて意味づけしないとしないとしたら、その行為は無意味である。そしてそれを行うのは、議会に対して責任を持つ市長だけではなく、教育長、そしてそこから依頼を受けた指導主事、そして学校長、さらに依頼を受ける研究主任、そして担任の先生。こうしたライン全体が機能させられる。まさに悲劇的な状況である。

前節の結果は、2007・2008年度に実施されたものを問題とした。今では他の形式となっている。表8に各年度に算数の問題として出題されたものの構成を、表9に示す略号によって記しておく。

表8：算数に関する「児童・生徒の学力向上を図るための調査」の構成

年\問	1	2	3	4	5	6	7	8	9	10
H28	技	知	知	技	考技	☆1	知考	☆2	☆3	☆3
H27	技	知	知	☆2	知	知考	考態	☆3	☆3	－
H26	技	知	知	☆2	知	知考	考態	☆3	☆3	－
H25	技	知	知技	知	考技	知態	知	考態	☆3	☆3
H24	技	知	考態	考技	技態	知	知	考	☆3	☆3
H23	技	知技	知	態	知	考	考技	知態	☆3	☆3

※「年」のH28は、2016（平成28）年度、H27は2015（平成27）年度、…に実施されたものを示す。「－」は、10問ではなく、9問の出題であったなどの理由で、当該の問いがないことを表す。

表9：表8における略号

略号	調査報告書での記述
技	数量や図形についての技能
知	数量や図形についての知識・理解
考	数学的な考え方
考技	小問（1）が数学的な考え方 小問（2）が数量や図形についての技能
知考	小問（1）が数量や図形についての知識・理解 小問（2）が数学的な考え方
態	算数への関心・意欲・態度
考態	小問（1）が数学的な考え方 小問（2）が 算数への関心・意欲・態度 (他, 「知態」など, 観点が2つの場合その順に記す)
☆1	小問（1）が数量や図形についての知識・理解 小問（2）が数学的な考え方 小問（3）が 算数への関心・意欲・態度 などのように3観点がこの順に。
☆2	考・技・態 の3観点がこの順に
取	必要な情報を正確に取り出す力
読	比較・関連付けて読み取る力
推	意図や背景, 理由を理解・解釈・推論して解決する力
☆3	(1) が, 「取」, (2) が「読」, (3) が「推」

これらの年度に関しては、それぞれ「算数への関心・意欲・態度」や、「読み解く力」も含めて、複数個の問いが為されている。内的妥当性の吟味を行うことは、不可能ではないが、報告書にはその結果に関する記述は見られない。せっかく同じ観点の問題を同一年度に複数出しているのであるから、相関が高いかどうか。そうした内的妥当性の吟味が求められるところである。教育測定論の教科書に書いてありそうなことではあるが、政策に関わる学力調査問題の作成者は、事後に妥当性の吟味を実際の試行結果を元に行うことで、調査問題作成のPDCAサイクルを完結させる義務がある。

なお、東京都がクロス集計を行っていないわけではない。例えば、平成28年度の調査報告書では、「調査の概要」に

「5 調査結果の公表

(1) 報告書による結果の公表

- ・各教科の観点ごと、読み解く力ごとの正答率（東京都全体）
- ・各教科の問題ごとの正答率（東京都全体）
- ・各教科の正答数分布（東京都全体及び各区市別の調査結果）
- ・児童・生徒質問紙調査集計結果（抽出校）
- ・学校質問紙調査集計結果（東京都全体 ただし、クロス集計は抽出校の調査結果による）

(2) 各教科の問題ごとの結果分析について

- ・抽出校の調査結果に基づき、分析した。

※抽出校とは、無作為抽出により選定した調査実施校の約 10%に当たる学校である」(p.4)

とあるように、[6] (3) 【結果分析】には、「正答の児童の教科全体の平均正答率は、誤答の児童より 10 ポイント以上高い。」(p.56) との記述がみられる。これは他の観点との関連を記述しているもので、この設問の妥当性を確認するための分析ではない。

地球の環境資源が有限であること。それを意識せざるを得ない時代となっている。トフラー (1980) が指摘するように、「第三の波」が押し寄せる必然であり、グローバル化が必要となるだろう。そして数値目標を立てて効率の良い行政が求められるところである。そこで、新自由主義者は賛同を集める。まさに歴史の流れの必然であったかもしれない。しかし、その「必然」が良いか悪いは別の問題だ。やっぱり教育とか福祉とかは聖域じゃないのか。そんな疑問や主張は大切にすべきだ。その数値が何を根拠にしてどのような意味を持つのかははっきりしない場合にでも数値でデータが出されると、正しいような錯覚が生ずる。新自由主義的な政策のもとでの数値を考える際に、特に警戒したいことである。

引用・参考文献：

- 東洋 (2001) 『子どもの能力と教育評価 (第 2 版)』 東京大学出版会 (UP 選書)
- 吉良直 (2009) 「どの子も置き去りにしない (NCLB) 法に関する研究—米国連邦教育法の制定背景と特殊性に着目して—」 日本教育大学院大学紀要『教育総合研究』 第 2 号, pp.55-71。
- 正田良 (2008a) 「国際調査による数学の評価枠組みの変容」 青山学院大学教育学会紀要『教育研究』 第 52 号 pp.85-97
- 正田良 (2008b) 『統計入門 因子分析の意味がわかる』 ベレ出版
- トフラー, A. (1980) / 徳岡孝夫訳 (1982) 『第三の波』 中公文庫
- 東京都教育委員会 (2008) 『平成 19 年度 児童・生徒の学力向上を図るための調査 報告書』
<http://www.kyoiku.metro.tokyo.jp/buka/shidou/chosa/pr080626g/19houkoku.htm>
- 東京都教育庁指導部義務教育特別支援教育指導課 (2010) 『平成 20 年度 児童・生徒の学力向上を図るための調査 報告書』 (東京都教育委員会印刷物登録 平成 21 年度 第 19 号)
- 東京都教育委員会 (2016) 『平成 28 年度 児童・生徒の学力向上を図るための調査 報告書』
http://www.metro.tokyo.jp/tosei/hodohappyo/press/2016/11/10/documents/18_02_01.pdf
- 渡辺治 (2013) 『安倍政権と日本政治の新段階—新自由主義・軍事大国化・改憲にどう対抗するか』 旬報社